Management Accounting Research, 2002, **13**, 229–248
doi: 10.1006/mare.2002.0180
Available online at http://www.idealibrary.com on IDEAL®

Integrated performance measurement design: insights from an application in aircraft maintenance

Paul Rouse*‡, Martin Putterill* and David Ryan†

This paper describes the development and application over a 4-year period of a performance monitoring system based upon mathematical programming methods. The setting is the engineering services division of an international airline. An integrated performance measurement system is developed that extends the balanced scorecard into an holistic appraisal system incorporating multiple perspectives and supporting measures. Data envelopment analysis (DEA) is used to quantify changes over time in productivity and continuous improvement. Fine-tuning of the DEA model to reflect the airline's strategy provides a balanced view of organizational performance. Further refinement and depth of analysis is obtained through the use of Malmquist techniques to trace the sources of performance change to efficiency and organizational learning. A noteworthy feature of the project was the very positive response from within the enterprise to the introduction and use of an application that provides appropriate and incisive performance feedback.

Key words: aircraft maintenance; balanced scorecard; continuous improvement; data envelopment analysis; malmquist; performance measurement; productivity.

1. Introduction

Intense competition and pressure to improve economic returns are forces driving international airlines to increase efficiency. One consequence has been the formation of alliances to integrate routings and to share the use of support facilities (Oum and

*Department of Accounting and Finance, The University of Auckland, Private Bag 92019, Auckland, New Zealand.
†Department of Engineering Science, The University of Auckland, Private Bag 92019, Auckland, New Zealand.
‡E-mail: p.rouse@auckland.ac.nz

Received 22 June 2000; accepted 7 November 2001.

1044–5005/02/$ - see front matter

Yu, 1998). To gain entry to an alliance, potential participants need to meet many criteria including the scope of existing networks and track record. Safety, efficiency and cost effectiveness are important determinants of suitability. Demonstrated ability to maintain aircraft at a high and consistent level of safe and efficient functioning is a crucial advantage.

Once part of an alliance even a relatively junior international airline partner can benefit from greater scale, scope and the chance to generate revenue from support work out-sourced by other participants. For airlines with the right level of skill this is an important strategic opportunity to establish and maintain competitive servicing capabilities within a multi-airline operating environment. One such opportunity is aircraft maintenance services. These service facilities must be capable of meeting all scheduled maintenance and intermittent emergency demands, necessitating effective integrated planning and control systems to ensure ongoing commitment to functional balance, responsiveness and operational improvement.

Over the past 20 years, management accounting has contributed to demands for more sophisticated systems through vigorous promotion of methods such as activity based costing (Cooper and Kaplan, 1991) and the balanced scorecard (Kaplan and Norton, 1992). Very significant developments in manufacturing of the early 1990s, notably the accent on quality, the value chain, and continuous improvement, were further stimuli to enhanced management accounting competence (Nanni *et al*., 1992; Shank and Govindarajan, 1992).

In order for these methods to deliver sustainable long-term improvements, they must be supported by performance measurement systems that can capture and report the critical time and complexity dimensions of particular enterprises. A call for greater sophistication in management accounting analytical tools and frameworks is made by Otley (1999, p. 375) who remarks 'mapping of means-ends relationships for a given organization is of crucial importance for the development of a meaningful Balanced Scorecard, and is worthy of much greater attention'. In a similar fashion, greater coherence concerning the linking of the goals and resources of each process to the overall goal of the company is also required (Nørreklit, 2000).

This paper describes the performance analytical tools and frameworks used to support change-management in the aircraft servicing and maintenance division (ASM) of an international airline (K). Data envelopment analysis (DEA) was used to measure productivity at departmental and divisional levels.[1] As the project developed, the importance of 'means-ends relationships' became apparent. A framework was developed that maps a close relationship between means-ends and comprises multiple perspectives in which measures reflect strategic directions and links to underlying resources and process drivers.

The integrated performance measurement system (IPMS) described was developed by a team consisting of members from KASM management and two of the authors. This research collaboration was undertaken over a period of 4 years from 1993 to 1997 and involved overlapping stages of theoretical assessment, design, field application and data analysis. To more effectively convey the phases and content of the project, this case study description follows a chronological format.

The KASM project describes the practical application of a number of contemporary management accounting innovations. Particular contributions are:

[1]This is the first description of an *ongoing* use of DEA over an extended period within an organisation.

- A description of an holistic framework for performance measures that enables the mapping of means-ends relationships;
- an extension of DEA beyond its usual cross-sectional analysis to a time-series application, which can be tracked over time as a measure of continuous improvement;
- the use of weight restrictions in DEA to provide a 'balanced view' of performance;
- a Malmquist decomposition to identify the impact of shifts in technology (in this case, organizational learning), separated from changes in productive efficiency.

The paper proceeds in Section 2 to introduce the business setting and change measurement imperatives. DEA, the cornerstone technique for productivity measurement, is described in Section 3, with results from the KASM field tests provided in Section 4. Details relating to the organisational response to its application in KASM and the development of a linked structure performance pyramid are outlined in Section 5. Section 6 describes the difficulties in obtaining a 'balanced view' and the solution adopted by KASM. Section 7 shows the part played in the airline maintenance setting of Malmquist analysis to determine the relative contribution to change of the effects on performance of learning and efficiency. Section 8 contains a summary and conclusion.

2. The organization and production models

KASM is a significant business unit with over 2000 full-time employees, maintaining approximately 150 domestic and international aircraft. KASM's stated mission is to become a 'world class' organization with a strong emphasis on productive efficiency in the field of aircraft maintenance. A high level of international regard for KASM's safety achievement indicated that this strategy has been effective.

For several years prior to engaging in this collaborative research, KASM had set about fostering a working environment conducive to continuous improvement. To consolidate and extend the overall productivity improvement goal, KASM was searching for a performance measurement system capable of influencing process outputs and outcomes,[2] and enabling superior performance to be identified and quantified to facilitate benchmarking. A motivating factor for the development of a new performance-monitoring tool was management's desire to improve the incentive scheme in KASM as part of initiatives to curb industrial problems.

Accordingly, there were several over-riding considerations in formulating the new KASM performance monitoring system. Of particular concern was the need to provide an investigative framework to extend performance assessment and achieve the 'coherence' referred to by Nørreklit (2000). Communication and interpretation of results in a form easily understood by all KASM employees were seen as equally important. This latter prerequisite resulted in the use of charts for most of the monthly reports. In addition, particular requirements for the system were articulated in two dimensions: (i) specific attention to the management of key areas

[2]Efficient and effective aircraft maintenance is vital for safety reasons, and is a crucial revenue driver since airlines earn revenue when planes are in the air, not when on the ground. Furthermore, the extremely high cost of replacement parts necessitates tradeoffs between inventory holding and stock-out costs.

of quality and inventory and (ii) an integrative capacity to synthesize and co-ordinate organizational goals, environment, structure and processes.

Within KASM responsibilities for aircraft maintenance service activities are undertaken by Department 1—Airframe, Department 2—Engines and Department 3—Components. Taking each Department in turn, principal tasks are (i) maintaining the overall aircraft, (ii) maintaining aircraft engines and (iii) storing and repairing components. Prior to the current study, KASM management had introduced a balanced scorecard to monitor productivity in each Department. The key measures used for gauging productivity were partial productivity ratios of output (chargeable hours) to individual inputs such as salaries, material usage and inventory, taking into account quality as measured by delivery performance. While management preferences focused primarily on these measures at the start of the prior phase, there were arguably other contenders for specific attention such as hangar occupation and equipment investment.

Intense collaborative effort in the first 18 months of the project resulted in the development and implementation of a new IPMS. Building on the prior scorecard experience, the initial focus of the research was on Departmental productivity measurement. The performance of each Department was measured separately each calendar month using the inputs and outputs described next.

Inputs

(1) Salary cost of productive time. This includes basic pay, overtime, penalties and allowances as well as idle time, but excludes leave and sick pay.

(2) Inventory cost. As shown below, this measure consists of two inventory-related figures and a set level for the proportion of each (alpha):

$$(1-\alpha) * Inv_{act}/n + \alpha * \{ABS(Mat_{act} - Mat_{fcst})\}$$

where Inv_{act} is the level of monthly inventory (dollar cost) averaged over the number of aircraft (n), ($Mat_{act} - Mat_{fcst}$) is a materials variance between actual and forecast usage,[3] and α is a proportion between zero and unity. The first item is of considerable concern to the airline due to the large carrying cost associated with the size and number of expensive spare parts.[4] Key management aims were to motivate Departments to reduce the overall level of inventories, while ensuring that stock-outs and consequent delays were minimized.

The second item targets materials usage performance. Departments forecast the expected cost of materials for each job. This forms the basis for comparison against actual costs with the variance being a measure of a Department's ability to minimize deviations from forecast levels of material usage.

During the pilot study, different levels were set for α. Finding after several trials that no useful insights were obtained, it was decided that each component should be weighted equally.

Outputs:

(1) Total hours charged, internally and externally plus time spent on capital improvements.

[3] *ABS* refers to the absolute difference regardless of 'under or over' usage.

[4] Each engine alone has an inventory value of US$5 million.

Departments carry out work for both their own and other airlines, and work on improvements to capital equipment. Although data for total chargeout revenue was available, hours charged were preferred as being more meaningful to operational staff.[5] Idle time is excluded for all Departments except Airframe, which has a very mixed range of activities. Under this arrangement, the cost of idle time is included in the salary cost input but not included as output except for airframe activities. This has the desired effect of penalizing Departments in months with significant idle time.

(2) Delivery performance.

A system was introduced that awarded positive and negative points for turn(around) time, and quality of service. Points on a reducing scale are awarded when an aircraft's maintenance is finished early, on time or is late. Delivery performance is rewarded where, in order to reduce passenger delays, aircraft arriving late are turned around sooner than the standard service time. Further points are awarded according to an assessment of how well customer requirements have been met i.e. a supplier's quality perspective. By scoring service efforts in this way, KASM management communicated the message that quality is an important attribute of performance.

It is clear from the foregoing that the KASM aircraft-servicing model is very focused and precise in intentions. Given the size and complexity of the division, the number of inputs and outputs in the model might seem overly small. This parsimony was in fact a conscious choice, reflecting the attitude of a management team determined in the first instance to change performance including employee behaviour, by taking simple and intelligible steps. There are parallel instances of a preference for the use of less 'accurate' measurement systems in the activity-based costing literature (e.g. Merchant and Shields, 1993). This KASM model formed the core of the sustained programme of departmental performance measurement and feedback described below.

3. Data envelopment analysis

The method used to analyse the monthly performance data is DEA, an approach first introduced by Charnes *et al*. (1978). The focus is on optimising an engineering-type ratio of outputs to inputs, by solving for a set of weights that satisfy a system of linear equations. The model (input-orientation) is as follows:[6]

$$\max z_0 = \frac{y_0^T u}{x_0^T v}$$
$$s.t.\ \frac{Yu}{Xv} \le 1$$
$$u, v \ge 0$$

[5]A further reason is that transfer prices for internal work are not based on market considerations.

[6]This fractional model is non-linear but as shown below can be transformed into a linear programming format in either multiplier (primal) or envelopment (dual) form:

	Multiplier		Envelopment
	$\max Z_o = \mathbf{y^T}\mu$		$\text{Min } h_o = \theta_o$
s.t.	$\mathbf{Y}\mu - \mathbf{X}\nu \le \mathbf{0}$	s.t.	$\mathbf{Y}\lambda \ge \mathbf{y_0}$
	$\mathbf{x^T}\nu = \mathbf{1}$		$\mathbf{X}\lambda \le \mathbf{x_0}\theta$
	$\mu, \nu \ge 0$		$\lambda \ge 0$; θ_o unrestricted.

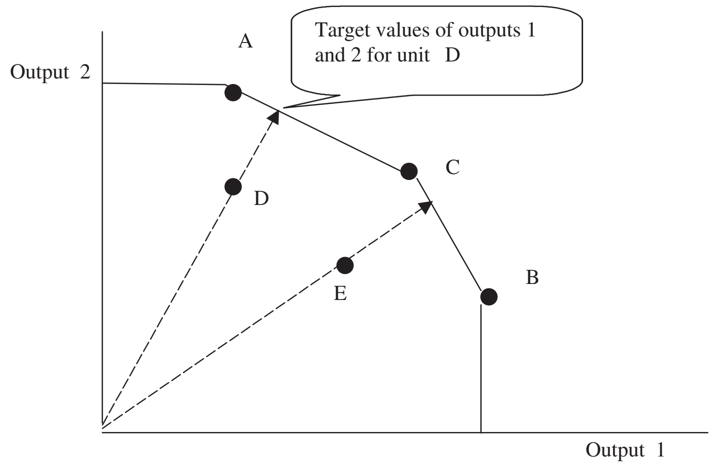

Figure 1. DEA efficiency frontier.

where $Y(nxs)$ and $X(nxm)$ are matrices containing s outputs and m inputs for n decision making units (DMUs), y_o and x_o are vectors of outputs and inputs for the particular DMU evaluated. The model searches for weights, u and v, which maximize the efficiency score for a unit 'o' subject to the constraint that these weights, when applied to the output–input ratios for all of the other units (1 to n) including unit 'o', do not result in an efficiency score of greater than 1 for any of these units.

Using a conventional cross-sectional example, consider five DMU's that produce two outputs using a single input. These DMUs could be departments, branches, firms or any specified areas of activity in which similar inputs and outputs prevail. To simplify, assume that they all use the same level of input but produce different amounts of the two outputs. Figure 1 shows that three of the DMU's produce different combinations of maximum amounts for the two outputs while DMU's D and E produce lesser amounts. DEA classifies A, B and C as fully efficient units that make up an efficiency frontier. D and E lie below this frontier and are deemed inefficient with efficiency scores measured by the distance along the ray extending from the origin through each unit to the frontier. D lies approximately 90% along its ray whereas E lies approximately 80%. Units on the frontier are assigned scores of 100% respectively. The efficiency scores for the five DMU's would be 100% for each of A, B and C and 90% for D and 80% for E.

DEA is an excellent tool for benchmarking. For example, not only does it provide an efficiency score for unit D (90%), but it also provides the target values that would make D efficient as shown in Figure 1. In addition, the units that are most suitable for D to benchmark itself against are units A and C, which form the segment of the efficiency frontier to which D is projected.

DEA models can be input or output oriented depending on whether inputs are to be reduced for inefficient DMUs while holding outputs constant, or outputs increased while holding inputs constant. DEA models can also be constant returns to scale or variable returns to scale, with minor modifications to the above model (see Banker *et al.*, 1984).

It should be noted that the only restriction on the weights is non-negativity.[7] It is therefore possible that an optimal solution for a DMU may include a weight of (near) zero for certain inputs and/or outputs. If managers regard this result as unsatisfactory, it may be desirable to impose upper and lower limits on the values that weights are permitted to take.

Imposing bounds on price ratios or virtual weights are two ways in which weights may be constrained. For example using price ratios and assuming two outputs with associated weights μ_1 and μ_2, the following relationship can be specified where L and U are lower and upper bounds respectively:

$$L \leq \frac{\mu 1}{\mu 2} \leq U.$$

This constraint can be interpreted as requiring the weight for output 1 to be 'valued' at least L times the weight for output 2 and at most, U times.

In most organizations, determining an appropriate price ratio is not an easy task. An alternative approach is to constrain the virtual weights by allowing managers to select lower and upper bounds for the contribution to efficiency a particular input or output is permitted to make:

$$L \leq \mu_1 y_1 \leq U.$$

For example, a manager may decide that output 1 should contribute at least 10% of the efficiency score for a DMU and at most 30%. In initial implementation stages, this is an easier task for managers than selecting price ratios.[8] In the KASM environment, management had no difficulty in setting bounds on the virtual weights of specific inputs and outputs, but were significantly challenged by the need to specify price ratio upper and lower bounds.

4. KASM field trial results

In KASM, the DEA analysis was applied to a situation where there were three Departments for which individual productivity measures were required. As previously indicated, measures of productivity were confined to inputs and outputs under the control of Department managers. Each monthly period is treated as a DMU within a single Department (i.e. performance of the same unit at different points in time).[9] As productivity was evaluated solely for each Department, constant returns to scale was agreed to be the appropriate model.

A carefully arranged system of data collection ensured that data from each Department were reliable and comparable. In particular, the financial data were adjusted where necessary for price increases (e.g. wage rate changes) and delivery performance reports were monitored closely by KASM management. Items such as

[7] To meet the requirement of strict 'positivity' in many conventional DEA models, the lower bound of zero is replaced with a non-Archimedean infinitesimal positive number.

[8] Once managers are satisfied with the virtual weight ranges, price ratios can be determined from examination of the efficient DMUs.

[9] Learning and efficiency implications of this time series application are provided in Section 7.

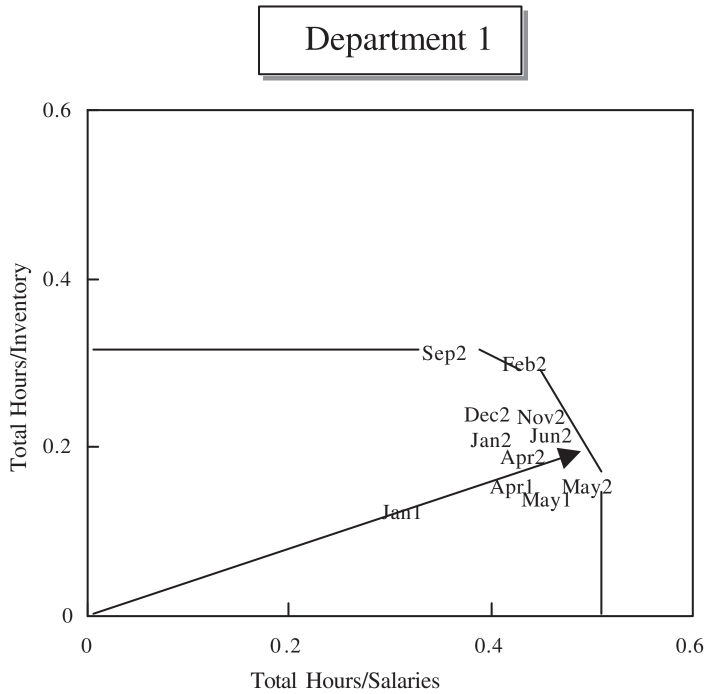

Figure 2. Ratio of total chargeable hours output to both inputs for KASM Department 1.

chargeable hours, inventory levels, materials usage and salaries were available from the conventional accounting system.

In order to clarify this application of DEA to a temporal as opposed to cross-sectional setting, Figure 2 shows a sample of months for Department 1 using performance over a 2-year period. The monthly ratios graphed are calculated by dividing one of the outputs, total chargeable hours, by the two inputs, inventory and salaries.

The months lying furthermost from the origin form an efficiency frontier from which the efficiencies of other months are calculated. In Figure 2, the most efficient months are Sep2 (being September of year 2), Feb2 and May2. Monthly performance is measured by its relative distance along a ray extending from the origin to the frontier. For example, Jan1 (January of year 1) is shown approximately two-thirds of the distance along the ray to the frontier, which would equate to an efficiency score of 67%. The use of graphs such as Figure 2 became a permanent feature for communicating the DEA measures of changes in performance between the company and engineering staff. Furthermore, in order to provide a Divisional context for Departmental performance, a division-wide measure was introduced.

Since the types of outputs and inputs are the same for each of the departments, surrogate division-wide measures can be obtained by summing the outputs and inputs across the three departments for use in calculating divisional DEA scores. This enables the productivity scores for a department to be reported relative to divisional performance, thereby removing any 'firm-wide' effects. Results are described as

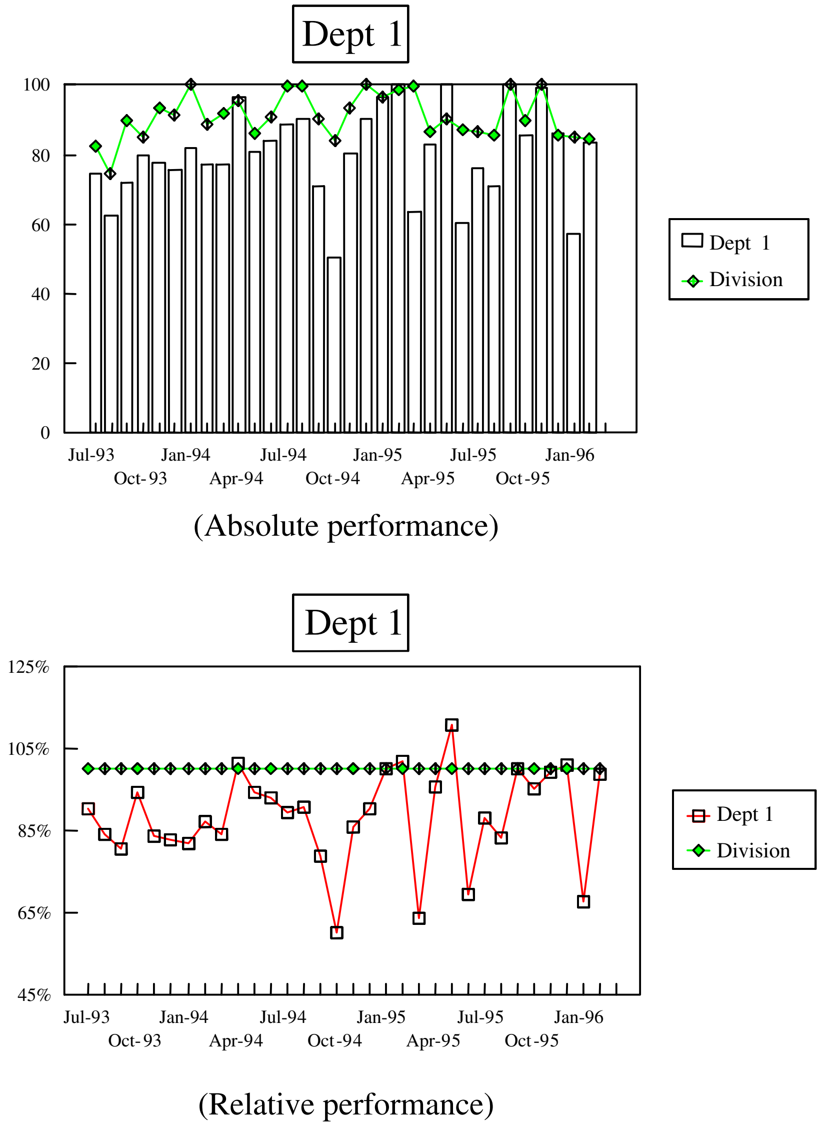

Figure 3. Absolute and relative DEA results for KASM Department 1.

'absolute' when departmental performance alone is considered, and 'relative' when departmental performance is related to the division.

Figure 3 shows graphs of DEA-derived measures for absolute and relative productivity for Department 1 between 1993 and 1996. The first graph, absolute Performance, highlights the most efficient months to have been Feb95, May95, Sep95, and Nov95. It is noted that although there is an increasing trend in the absolute efficiency scores up to July94, some severe fluctuations in monthly performance thereafter offset the excellent results in the 4 months mentioned.

The lower graph (Figure 3) shows relative monthly DEA results for Department 1 against KASM as a combined division. An improvement is seen in isolated months in

1995 but performance is still uneven. A possible reason for this unevenness is the use of calendar months as opposed to maintenance cycles, e.g. a major check and service for an engine may span several months.

Apart from these considerations, KASM personnel responded favourably to DEA output in this form, finding results easy to understand. A key factor in the climate of acceptance of the new system of performance reporting was the provision of training and interpretation of results. In addition to these composite DEA scores, each month staff were given supporting information such as target values for improvement, and a listing of 'efficient' DMUS, i.e. months on the efficiency frontier. A regular system of reporting DEA results was established within each department based on similar graphs to those shown above.

As the study progressed, several factors began to trigger management concerns that the performance measures were not appropriately 'balanced' from a *technical DEA perspective*. First, the initial flexibility of weight selection in DEA resulted in measures considered to be important components of performance, being ignored through the assignment of zero weights to them. Second, the weightings on some measures were thought to be disproportionate to their perceived importance. As quality and inventory management were part of the division's strategic goals, upper and lower bounds were placed on the virtual weights for delivery performance, and lower bounds on the virtual weights for the two inputs.

In the case of delivery performance, a policy was introduced that at least 10% (with an upper limit of 30%) of the efficiency score for a DMU (a month) would be from this output. In effect, an efficient month would be required to have at least 10% and at most 30% of the 100% efficiency score attributed to delivery performance. The effect of these bounds is to limit the virtual weight for combined hours to a maximum 90%. For inputs, a lower bound of 20% was set.

Shortly after the new measurement system was put into operation, concern also surfaced about the interpretation of results. While monthly measures of productivity, target values and 'best performance' months were routinely reported for each department, substantive reasons for performance variability were not always obvious. For example, although the main component of poor performance for a month might be an unfavourable materials variance decision-makers still lacked incisive information on underlying causes. The synoptic perspective of the financial reporting system was patently too blunt for focusing intervention.

In response to the need for more incisive explanations of the DEA output, a high-level model of KASM operations was developed in the performance framework described next.

5. Performance frameworks

As the new performance measurement became better understood and accepted, KASM departmental managers and employees requested better information about possible causes of changes in monthly results. This call for greater depth of analysis was a positive endorsement, but in order to move to the coherence described by Nørreklit (2000) it would not be enough simply to provide sharper indicators; some clear sign-posting of directions for specific as well as holistic action is essential when raising the complexity and sophistication of performance measurement systems.

This is not an easy task since when formulating these systems, several interrelated issues appear to be barriers to their development as effective performance appraisal practices.

'Unboundedness', e.g. the propensity for lists of measures to increase when there are no bounds imposed. Meyer (1998, p. xvi) reports 50 to 60 measures are commonplace in many firms and cites the case of one firm that uses 117 'top-level' measures.

'Lack of context' refers to interpretation difficulties when measures are considered in isolation. Arguably there is a need to relate performance to a setting, i.e. for meaningful appraisal of results, contextual information is an essential requirement when measuring significance and relevance.

'Incompleteness' is one of the lynch pins justifying the multiple perspective approach of the balanced scorecard (Kaplan and Norton, 1992). It recognizes not only a need for financial indicators but also non-financial measures to accommodate the diverse management tasks of different stakeholders.

'Behavioural' problems leading to dysfunctionality. For example, too much emphasis on an 'indicator' may lead employees to ignore the underlying processes and concentrate on measurements. Taken to the extreme, this selection/direction process opens the way for the measurement process to become political with dysfunctional instances of top–down versus bottom–up selection, pseudo-participation processes, and weakened goal congruence.

Mindful of these issues, a performance framework was developed from a number of earlier proposals to address performance issues. These include (i) Anthony (1965) framework for planning and control systems, (ii) the performance pyramid of Lynch and Cross (1991) that provides a representation of the linkages between strategic vision and operations; (iii) the need identified by Beischel and Smith (1991) for performance measures to link critical success factors (CSF) to process levels, and (iv) the most widely promoted structure consisting of a selection of multi-dimensional perspectives in the form of a balanced scorecard (Kaplan and Norton, 1992) that pays explicit attention to the need for relevant information about financial stakeholders, customers, internal management and innovation and learning.

In Figure 4, each face of the pyramid reflects one of the four perspectives of the balanced scorecard.

The CSF shown in Figure 4 embody the strategies chosen to move the enterprise in the direction of the organization vision. Within each face, performance (managerial) measures were sought to link the underlying process drivers to the CSF in the manner suggested by Beischel and Smith (1991). Measures and process drivers are not only linked upon each face of the pyramid but connections also exist to other faces. This enables managers to understand the impact of process drivers on more than one key result area, e.g. productivity and quality (Maani *et al.*, 1994).

From this conceptual starting point, performance measures can be chosen that in aggregate are capable of presenting a comprehensive, integrated and activity-related view of a complex business situation. Specifically, the IPMS construct ensures that:

(i) Performance measures related to CSF are explicitly linked to one or more elements of strategy (i.e. the faces of the pyramid).

(ii) The linked structure and performance measures when determined, are essential components in any further network or path analysis aimed at improving or modifying linkages.

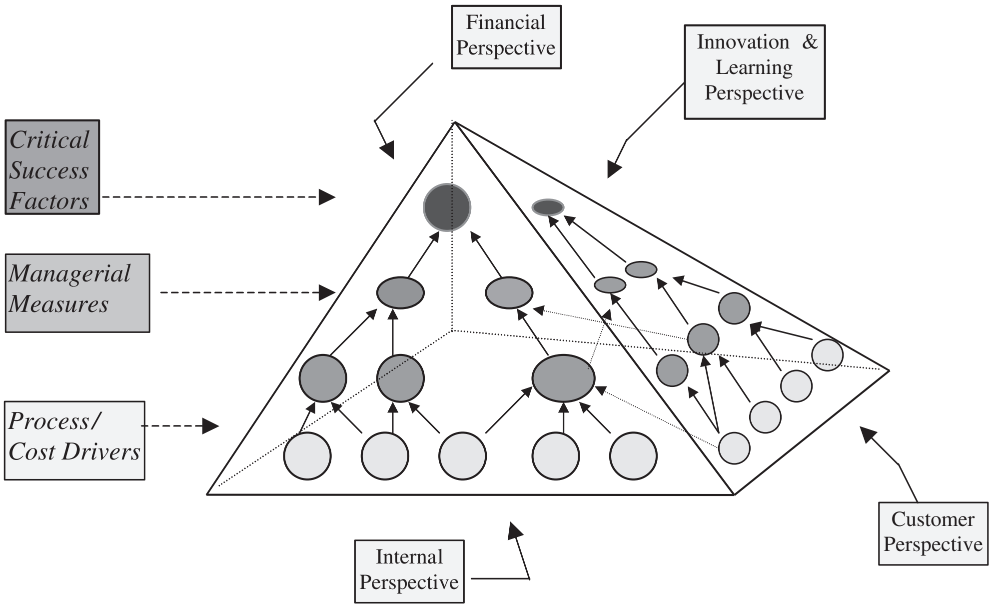

Figure 4. IPMS incorporating the balanced scorecard, performance pyramid and linked measures.

(iii) Signposts are provided to underlying causes via the linkages between measures and process or cost drivers;

(iv) Through its use, the mission and strategy are effectively communicated throughout the organization.

KASM IPMS pyramid

With a KASM pyramid to direct attention, CSFs, together with performance data cascading down to underlying drivers were identified for each perspective (i.e. each face of the pyramid). The CSFs chosen for each perspective were (i) quality—for the customer focus, (ii) productivity—for the internal focus, (iii) economic value added—for the financial focus and (iv) organizational learning—for the innovation and learning focus. Each of these was in turn broken into components. For example, quality was defined as (a) meeting or exceeding customer needs and (b) achieving customer satisfaction. These two quality considerations can be measured by delivery performance and on-line-on-time performance, respectively. An operational sub-measure such as turn time, e.g. time to service an aircraft and have it back in operation, can be further assessed by reference to timeliness and rework measures, and associated process drivers e.g. equipment availability, professional skills and so on.

For the internal perspective, productivity was defined as optimising outputs to inputs in inventory, fixed capital and people. Each is represented by labour hours charged, stock turnover, facility or hangar utilization, and salary costs. Using inventory by way of illustration, stock turn is supported by inventory level and material usage. Underlying drivers for inventory level are safety stock policy, vendor lead-time and price changes. An interesting driver for material usage is average flight sector length, which is determined by operational decisions on flight scheduling.

The shorter the average sector length, the greater the frequency of take-off and landings; hence the greater wear and tear on landing gear and tyres.

As the new performance measurement system began to generate results, several important insights emerged. First, it soon became clear that certain drivers of maintenance productivity are triggered by decisions made outside KASM, e.g. average sector length. Second, once it was realized that facility availability was regularly being identified by DEA as a significant explanatory factor, managers began to give commensurate attention to this dimension of the enterprise. Third, the importance of organization learning through human resource management could be clearly seen by the links to delivery performance and quality. Fourth, management very soon concluded that communication with staff would be substantially improved by relating the charts of DEA performance scores to the visual representations of the IPMS pyramid. Each month DEA results were circulated and likely causes identified using IPMS.[10] An emphasis on graphical reporting played a key role in fostering greater interest by staff in relationships among processes and activities and in the impact that lower-level problems have on higher level measures and CSF.

DEA provided composite efficiency scores and benchmarking information that greatly facilitated testing of alternative measures and enabled the project to proceed at a rapid pace. After initial debate and experimentation in the choice of measures, a DEA model capable of capturing movements in productivity in an accurate and timely fashion was accepted by KASM management.

6. Holism, myopia and balance

The KASM framework enabled decision-makers in the division to incorporate their detailed knowledge of policy, process and purpose into the interpretation of performance and consequent continuous improvement interventions. This analysis and reaction sequence involving staff at all levels is a far more active approach to fostering continuous improvement and the strategic dialogue proposed by Nørreklit (2000), than other ways to assemble performance indicators. This includes the widely disseminated balanced score card approach to informing decision-makers by means of an assortment of measures in four segments of supposedly equally importance.

Particular characteristics of the KASM performance measurement system, including the supportive role of DEA, warrant special consideration. A crucial distinguishing characteristic of the approach is the channelling of efforts and resources to areas of need and monitoring consequences in an holistic way in full recognition of the linked structures.

Nonetheless, as with any framework comprising a number of measures and multiple perspectives, management must be alert to any tendencies for the division to drift out of control for reasons that include the following.

(i) There are cognitive limitations upon individual's information processing abilities. 'Because individuals have limited powers of understanding and can deal with only very small amounts of information at a time, they

[10]To provide emphasis to these causal explanations, a specific process drivers was colour-coded each time it was found to be a factor explaining the DEA results. This provided a powerful visual pinpointing of underlying problems.

inevitably display limited rationality' (Emmanuel *et al.*, 1990, p. 49). Not only are individuals unable to cope with more than a small, limited set of information, Macintosh (1985, p. 46) describes how beyond a certain point, further information serves only to fossilize judgements due to overconfidence on the part of the expert judges.

(ii) Fossilization and overconfidence contribute to functional fixation manifested by a bias towards particular measures or perspectives. Within an organization, it is easy to contemplate financial accountants focused on financial measures, production personnel on internal perspective measures, marketing staff on customer perspectives, and research and human resource personnel on innovation and learning perspectives. Although wider perspectives may be encouraged, there appears to be a natural tendency to give attention to the familiar.

(iii) The notion of 'organisational balance' is poorly articulated even in seminal papers on performance appraisal, with the result that managers in general lack guidance as to what is meant by 'balance'. For example, Atkinson *et al.* (1997, p. 532) compound rather than resolve the issue of interpreting balance in the balanced scorecard. By requiring performance measurement systems to both reflect the organization's understanding of causes of successful performance on its primary goals (the depth requirement), and measure the most critical aspects or differentiators of organizational performance (the breadth requirement), they appear to be relying on an inoperable duality. Yet even if this distinction could be made, these requirements do not overcome problems of cognitive limitations, fossilization and functional fixation.

In the KASM study, a special effort was made to safeguard against myopic behaviour. A key feature of this effort was the imposition of lower and upper bounds on virtual weights within the DEA model.[11] Preferences are made explicit while preserving degrees of flexibility; a feature lacking when uniform weights are adopted[12] or implied (as in the balanced scorecard). Technically, if a feasible solution cannot be obtained, conflicting managerial preferences are immediately identified. In this manner, DEA in particular provided the means for preserving organizational holism and avoiding imbalance due to myopia.

7. Learning effects and technology change

As the KASM performance appraisal system moved into an operational phase, the academic involvement became more reflective. Additional dimensions of performance were explored at a more refined technical level that involved further decomposition to separate out changes due to doing things better (efficiency), and those due to better technology. The use of KASM time series data in the form of DMUs differs from conventional DEA cross-sectional applications. Consequently, there is an uncertain combination of learning and efficiency effects. The fact that

[11] Smith (1997, p. 35) suggests a similar solution to the problem of balance between 'financial results and operational measures of internal processes, innovation and improvement activities. Managers need to be able to view performance across several dimensions simultaneously—dictating a multivariate approach.'

[12] A similar notion can be found in the use of fuzzy logic sets as described in Rangone (1997).

Table 1
Malmquist Index for KASM showing changes in efficiency and learning (represented by shifts in the frontier)

	Change in efficiency (1) (%)	Shift in frontier (learning) (2) (%)	Combined Malmquist (1) ∗ (2) (%)
January	100	92	92
February	116	112	129
March	106	111	117
April	86	95	82
May	112	113	127
June	109	108	117
July	76	109	84
August	72	108	78
September	100	112	112
October	109	112	122
November	109	111	121
December	90	108	97
Means	98.75	107.58	106.50

there is a mixture makes it difficult to identify when performance improvement reflects a change in performance efficiency and when it reflects learning. A solution to this is offered by using a Malmquist (1953)[13] decomposition that separates shifts in the frontier from changes in efficiency.

A shift in the frontier is interpreted as a shift in the rate of learning. In other words, managers improve both their learning processes (changes in efficiency) as well as their learning systems (shifts in the frontier). A similar effect is described by Lapré *et al*. (1999) who distinguish between *operational* learning ('the process of obtaining validation of action-outcome links') and *conceptual* learning ('the process of acquiring a better understanding of cause-and-effect relationships').

Figure 5 conceptually sketches a two output, single input example for two periods, year $1(t)$ and year $2(t + 1)$, for a constant returns-to-scale technology. In order to simplify the explanation, performances for a single month (September) in years 1 and 2 are shown.

In Figure 5(a), Sept2 (S2) lies on the frontier and is fully efficient. Sept1 (S1) is below the frontier and is 96% efficient. The changes between Sept1 and Sept2 can be decomposed into changes due to efficiency and to technology by constructing a frontier for each year and evaluating S1 and S2 against the year 1 and year 2 frontiers. Using year 1 technology only (ie excluding year 2 months), S1 is fully efficient and forms part of the frontier in Figure 5(b). S2 is fully efficient using year 2 technology (Figure 5(a)) and therefore, each September month is efficient relative to the technology existing within their respective years. Figure 5(c) combines (a) and (b) showing: (1) September lying on the frontier for each year's technology, and (2) that the frontier has shifted outwards between years 1 and 2. It is pertinent to note that year 2 is not always better than year 1. This situation is illustrated in Figure 5(d) where the old technology is superior to the new technology along particular segments of the frontiers (i.e. where output 1 is favoured).

Against this conceptual background, Malmquist methodology was applied to two calendar years of actual KASM data. Table 1 reports the monthly Malmquist results.

[13]See Appendix for a more detailed description of the Malmquist index.

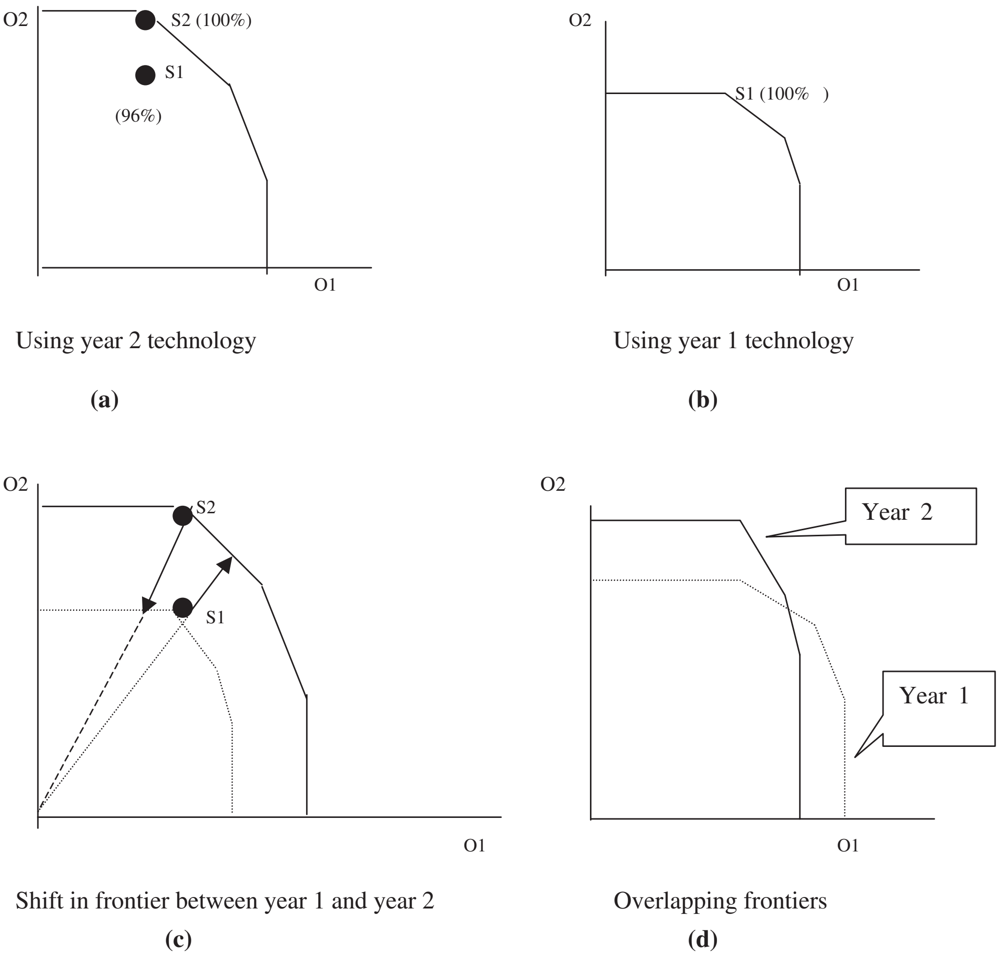

Figure 5. Year-on-year single month changes in departmental efficiency and technology using Malmquist and two outputs (O1 and O2).

The first column shows the change in efficiency between the 2 years and the second column reports the shift in the frontier. The product of these first two columns is shown in the third column being the combined Malmquist Index. In the manner alluded to in Figure 5, Table 1 results show that while September remained an efficient month in both years, the frontier has shifted outwards representing an improvement in technology. July and August's relative efficiencies have worsened but are compensated by an outward shift in the frontier (interpreted as learning in this study). This form of interpretation raises several difficult issues e.g. note how January and April reflect a contraction of the frontier.[14]

The results in Table 1 suggest that overall, gains have been achieved in efficiency and learning culminating in across the board improvement. Learning is a major

[14]These could have been 'bad' months when things did not go well. Interestingly, in the Southern Hemisphere these months coincide with summer and Easter holidays.

contributor to improvement and has lifted the level of performance on average by 8% (compared with 1% average decline in efficiency).

While more incisive information concerning the sources of improved performance has potential management uses in areas such as incentive schemes and investment strategy, this analysis remained in the academic domain for several reasons. First, Malmquist analysis is still in the exploratory stages within the DEA research arena and the interpretation of frontier shifts as lifting the learning platform is novel.[15] Second, it is doubtful whether at such a formative stage in building commitment, KASM would have wanted an analysis showing that performance had been raised almost entirely due to management initiatives rather than improvements in efficiency. Third, the greater level of complexity of the Malmquist interpretation would almost certainly have been an obstacle to its acceptance as a practical tool in this setting at this particular stage.

This experiment and Malmquist analysis has revealed a potential to become an important tool for productivity analysis, and strategic management. Porter (1996) refers to a productivity frontier of operational effectiveness that shifts in response to competitive intensities. Maintaining relative efficiency whilst keeping pace with shifts in technology is a challenge for organizations. Information provided by the Malmquist technique appears tailor-made for this purpose.

8. Conclusion

The holistic performance framework supported by DEA, developed through private firm and academic collaboration and field application, was successfully implemented in this aircraft maintenance setting. The extent of progress and achievement of this innovative management accounting approach is evidenced by reports such as those shown in Figure 3, which formed a significant part of the regular performance reporting system. The linked structure performance pyramid was recognized by the airline as a major management tool with charts tailored to individual departments being regularly posted on departmental notice boards. Throughout its use over a 4-year period, DEA efficiency scores were the major measures of productivity and were used as reference points for evaluating the divisional incentive schemes.

The system described appears to have effectively met KASM needs. The main challenges were to build a clear picture of operational linkages, combine these elements 'holistically', and to select realistic data to reflect the important control characteristics. The collaboration between the airline and university personnel was very reliant on the commitment of the particular KASM representatives who championed the innovation. Not only did they make significant technical contributions, but they were active in initial internal education of staff in the use and interpretation of results, as well as in data gathering and quality assurance.

A major lesson learned from this study was that while methods such as DEA provide the 'bones' of performance analysis, the measurement structure provides the 'body' for successful performance evaluation and measurement. As shown, the pyramid structure was developed several months after the DEA study had commenced

[15] A detailed treatment of Malmquist in temporal and network scenarios is provided in Färe and Grosskopf (1996).

in response to a need to understand and act upon the DEA results. A further critical observation is that although a balanced scorecard had been developed beforehand, it did not enable managers to drill down to underlying drivers or to comprehend the interrelationships among strategic goals, measures and drivers. Furthermore, the use of weight restrictions employed by the DEA model show how managers can increase the relevance of performance measures by judicial 'balancing' decisions.

The Malmquist method provides potentially valuable insights into efficiency and the effects of learning. The airline had instigated deliberate processes directed towards continuous improvement and management believed that these efforts had been successful. The DEA results confirmed these beliefs and the subsequent Malmquist decompositions enable the nature of the improvements to be identified. The ability to decompose productivity change into learning and efficiency components fit well with efforts organised around target costing and kaizen (Lee *et al*., 1994). Strategically, improvements due to learning initiative must be matched with concurrent improvements in efficiency to capture the opportunities arising as the frontier expands. Otherwise new territory created by 'pioneers' can fail to be developed by lethargic 'settlers'.

Two areas are suggested for further research. First, the relationship between DEA measures of productivity and financial results needs to be identified and reconciled for differences in timing and measurement (e.g. the link between improved delivery performance and increased cash flows). Second, translating performance measures into appropriate managerial action requires knowledge of the underlying causal process drivers. The directed network structure of the performance pyramid suggests potentially valuable applications for structural causal modelling of the relationship between drivers and performance.

Given growing recognition that fostering and measuring continuous improvement are cardinal management accounting attributes, greater sophistication in performance representational frameworks and analytical tools is not only appropriate for the matter at hand, but is a signpost of the direction of advance in value-adding capability for the discipline.

Acknowledgements

We would like to thank Tae Oum, University of British Columbia, and Shane Moriarity and William Maguire from the University of Auckland for their assistance. We also thank the Editor Robert Scapens and an anonymous referee for their helpful comments.

Appendix

Malmquist index

Using the Malmquist output indices developed by Caves *et al*. (1982), Färe *et al*. (1989) defined the following Malmquist-type measure of productivity:

$$M_o(x^{t+1}, y^{t+1}, x^t, y^t) = \sqrt{\left[\frac{D_o^t(x^{t+1}, y^{t+1}) D_o^{t+1}(x^{t+1}, y^{t+1})}{D_o^t(x^t, y^t) D_o^{t+1}(x^t, y^t)}\right]}$$

with an equivalent form:

$$M_o(x^{t+1}, y^{t+1}, x^t, y^t) = \frac{D_o^{t+1}(x^{t+1}, y^{t+1})}{D_o^t(x^t, y^t)} \sqrt{\left[\frac{D_o^t(x^{t+1}, y^{t+1}) D_o^t(x^t, y^t)}{D_o^{t+1}(x^{t+1}, y^{t+1}) D_o^{t+1}(x^t, y^t)}\right]}$$

where $D_o^t(x^t, y^t)$ = efficiency or distance measure for a unit o using period t technology and period t inputs and outputs; $D_o^{t+1}(x^{t+1}, y^{t+1})$ = efficiency or distance measure for a unit o using period $t+1$ technology and period $t+1$ inputs and outputs; $D_o^{t+1}(x^t, y^t)$ = cross-efficiency or distance measure for a unit o using period $t+1$ technology and period t inputs and outputs; $D_o^t(x^{t+1}, y^{t+1})$ = cross-efficiency or distance measure for a unit o using period t technology and period $t+1$ inputs and outputs.

This approach enables productivity measures to be decomposed into efficiency and technical changes corresponding to the first and second terms respectively (the second term is enclosed by the square brackets). 'The efficiency change term is equivalent to the ratio of Farrell technical efficiency in period $t+1$ divided by [the] Farrell technical efficiency in period t. The technical change term is the geometric mean of the shift in technology as observed at x^{t+1} (the first ratio inside the bracket) and the shift in technology observed at x^t (the second ratio inside the bracket)' (Grosskopf, 1993, p. 177).

The 'cross-period efficiencies' can be calculated using the modified DEA models described in Andersen and Petersen (1993), Banker *et al*. (1989) and Färe and Grosskopf (1996). Briefly, these are calculated by excluding the DMU under evaluation (conventionally denoted DMU_0) from the technology reference set. In the envelopment model, this is equivalent to:

$$\begin{aligned} \mathrm{Min} h_o &= \theta_o \\ Y\lambda_j &\geq \mathbf{y_o};\ \mathbf{o} \not\subset \mathbf{j} \\ \mathbf{X}\lambda_\mathbf{j} &\leq \mathbf{x_o}\theta;\ \mathbf{o} \not\subset \mathbf{j} \\ \lambda &\geq 0;\ \theta_o \text{unrestricted.} \end{aligned}$$

References

Andersen, P. and Petersen, N. C. A, 1993. A Procedure for ranking efficiency units in data envelopment analysis, *Management Science*, **39**, 1261–1264.

Anthony, R. N., 1965. *Planning and Control Systems: A Framework for Analysis*, Boston, Graduate School of Business Administration, Harvard University.

Atkinson, A. A., Banker, R. D., Kaplan, R. S. and Young, S. M., 1997. *Management Accounting*, 2nd edition, Prentice-Hall.

Banker, R. D., Charnes, A. and Cooper, W. W., 1984. Some models for estimating technical and scale efficiencies in data envelopment analysis, *Management Science*, **30**, 1078–1092.

Banker, R. D., Das, S. and Datar, S. M., 1989. Analysis of cost variances for management control in hospitals, *Research in Governmental and Nonprofit Accounting*, **5**, 269–271.

Beischel, M. E. and Smith, R., 1991. Linking the shop floor to the top floor, *Management Accounting*, October, 25–29.

Caves, D., Christensen, L. and Diewert, W. E., 1982. The economic theory of index numbers and the measurement of input, output and productivity, *Econometrica*, **50**, 1393–1414.

Charnes, A., Cooper, W. W. and Rhodes, E., 1978. Measuring the efficiency of decision making units, *European Journal of Operational Research*, **2**, 429–444.

Cooper, R. and Kaplan, R., 1991. *The Design of Cost Management Systems: Text, Cases and Readings*, Prentice-Hall.

Emmanuel, C., Otley, D. and Merchant, K., 1990. *Accounting for Management Control*, Chapman and Hall.

Färe, R and Grosskopf, S, 1996. *Intertemporal Production Frontiers: with Dynamic DEA*, Kluwer Academic Publishers.

Färe, R., Grosskopf, S., Lindgren, B. and Roos, P., 1994. Productivity developments in Swedish hospitals: a Malmquist output index approach, Chapter 13, in A. Charnes, W. W. Cooper, A. Y. Lewin, L. M. Seiford (eds), *Data Envelopment Analysis: Theory, Methodology and Applications*, Kluwer Academic Publishers.

Grosskopf, S, 1993. Efficiency and productivity, Chapter 4, in H. O. Fried, C. A. K. Lovell, S. S. Schmidt (eds), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press.

Kaplan, R. S. and Norton, D. P., 1992. The balanced scorecard—measures that drive performance, *Harvard Business Review,* January–February, 71–79.

Lapré, M. A., Muhkerjee, A. S. and Van Wassenhove, L. N., 1999. Behind the learning curve: linking learning activities to waste reduction, *Management Science*, **46**, 587–611.

Lee, J. Y., Jacob, R. and Ullinski, M., 1994. Activity-based costing and Japanese cost management techniques: a comparison, *Advances in Management Accounting*, **3**, 176–196.

Lynch, R. L. and Cross, K. F., 1991. *Measure Up!*, London, Mandarin.

Maani, K. E., Putterill, M. S. and Sluti, D. G., 1994. Empirical analysis of quality improvement in manufacturing, *International Journal of Quality & Reliability Management*, **11**, 19–37.

Macintosh, N. B., 1985. *The Social Software of Accounting and Information Systems*, John Wiley & Sons.

Malmquist, S., 1953. Index numbers and indifference surfaces, *Trabajos de Estadistica*, **4**, 209–242.

Merchant, K. A. and Shields, M. D., 1993. When and why to measure costs less accurately to improve decision making, *Accounting Horizons*, **7**, 76–81.

Meyer, M. W., 1998. Measuring and managing performance: the new discipline in management, in A. Neely, D. B. Waggoner (eds), *Performance Measurement—Theory and Practice*, Centre for Business Performance, University of Cambridge, UK.

Nanni, A. J., Dixon, J. R. and Vollmann, T. E., 1992. Integrated performance measurement: management accounting to support the new manufacturing realities, *Journal of Management Accounting Research*, **4**, 1–19.

Nørreklit, H., 2000. The balance on the balanced scorecard—a critical analysis of some of its assumptions, *Management Accounting Research*, **11**, 65–88.

Otley, D., 1999. Performance management: a framework for management control systems research, *Management Accounting Research*, **10**, 363–382.

Oum, T. H. and Yu, C., 1998. Winning airlines: productivity and cost competitiveness of the world's major airlines, *Transportation Research, Economics and Policy*, **6**, xvi, 220.

Porter, M. E., 1996. What is strategy? *Harvard Business Review,* November–December, 61–78.

Rangone, A., 1997. Linking organizational effectiveness, key success factors and performance measures: an analytical framework, *Management Accounting Research*, **8**, 207–219.

Shank, J. K. and Govindarajan, V., 1992. Strategic cost management and the value chain, *Journal of Cost Management,* Winter, 5–21.

Smith, M., 1997. Putting NFIs to work in a balanced scorecard environment, *Management Accounting* (UK), March, 32–35.